DOCUMENT RESUME

ED 224 350 FL 013 401

AUTHOR McKee, Macey Blackburn
TITLE Sentence-Combining -- Not If or When, but How.
PUB DATE 82
NOTE 11p.; In: Boyd, John R., Ed.; Haskell, John F., Ed. "Selected Papers from the Illinois TESOL/BE Annual Convention" (10th, Chicago, IL, February 26-27, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Bound Students; Discussion (Teaching Technique); *English (Second Language); *Error Analysis (Language); Instructional Innovation; Language Research; Linguistic Competence; Questioning Techniques; Secondary Education; Second Language Instruction; *Sentence Combining; Sentence Structure; *Syntax; *Writing Exercises

ABSTRACT

A method of teaching English as a second language (ESL) involving open sentence combining practice with group discussion was compared with a more traditional teaching method. Students were largely college-bound and attended the University of Arizona's ESL Center. Sentence combining was analyzed for a paragraph-length exercise on a low-level scientific topic and a free-writing sample on a daily conversation topic, which were produced in the first and last weeks of an 8-week session. For 7 weeks, the traditional instruction (control) group completed text exercises and some compositions and identified and corrected errors as a class. During this same time span, the experimental group completed a free-writing assignment focusing on some particular structure or an open-writing exercise in sentence combining using a variety of structures. They were also given an essay assignment designed to elicit a given structure. It was found that the type of sentence combining practiced by the experimental group resulted in increased syntactic maturity and a decrease of error. The control group sustained losses, particularly in the error-free measures as they attempted to use the complex structures that they had studied but not mastered. Attention is directed to types of errors, preliminary findings of a 16-week study, and a hypothesis concerning the relationship between increased syntactic maturity scores and improvement in reading comprehension. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ILLINOIS TESOL/BE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED224350

SENTENCE - COMBINING -- NOT IF OR WHEN, BUT HOW

Macey Blackburn McKee
University of Arizona

"Saying at this juncture that sentence combining needs more study is like telling Thomas Edison his light bulb needed more work just after he'd thrown the switch illuminating the street lights of New York." (Mellon, 1979)

In my opinion, research on sentence combining is just beginning. What had been empirically established when this was said was that more students in some kind of sentence combining group had showed more improvement in syntactic maturity as measured by some T-Unit index (Hunt, 1965) than had the controls in some kind of "traditional" program; that there was reason to believe that this type of syntactic maturity score correlated to some extent with holistic ratings; and that the question of the nature of the relationship between syntactic maturity scores and reading comprehension was ambiguous. This description offers a number of lines of research needed even in native speaker work to settle all the questions that even true believers have. Far less has been done in FL research (Cooper, 1976, 1978, 1980) and very little in ESL. The work to be reported here will touch upon all of these lines of inquiry. It is research in progress, with only two classes in each of the three projects, so no definitive claims will be made; however, most of the statistical results of these formal studies bear out the informal observations made over a period of years.

Zamel (1980), in reviewing selected literature, points out the difficulty of ascertaining from the published report just exactly what took place, especially in the control group, often described as having followed the "traditional" curriculum. It is sometimes equally difficult to determine exactly what took place in the experimental condition; for example, many reviewers have begun with Bateman-Zidonis (1966) although there was no sentence combining work in that study. Since these studies are expressly concerned with the "how", the methods and the types of materials, it has been a constant source of frustration to find so little information in many reports about these matters. There is no reason to expect all the many varieties of sentence combining to have similar effects, especially as compared with a multitude of "traditional" courses.

FL013401

This research is taking place in the most advanced level of the Center for English as a Second Language at the University of Arizona. The students involved have actual or assumed TOEFL scores ranging from 400 to 650 at entry into this level. Between 95% and 100% of a given session's population is university-bound, many for graduate school. The courses at this level are ESPs of various natures; thus, random assignment is not possible. The laboratory course for these groups is the same; all others are different except the writing class. The writing classes are grammar classes, with the explicit goal of error-free syntactic maturity, all using the same text (Frank, 1972) if a text is used, supplemented by whatever each instructor chooses to add.

From John R. Boyd and John F. Haskell, eds., "Selected Papers from the Tenth Illinois TESOL/BE Annual Convention" (Chicago, IL, Feb 26-27, 1982)

We are now in the third project. This discussion will center primarily on last summer's 8-week project because the data from the subsequent 16-week project are not yet completely analyzed.

There were initially two broad questions: 1) Does "my" type of sentence combining produce greater improvement in error-free syntactic maturity than does our traditional writing class? 2) Is there a relationship between increased syntactic maturity scores and improvemen in reading comprehension? There are many more specific questions, some of which I will mention later.

With these questions in mind, I collected date from the three sections that existed during the summer. After inspecting the writing samples, I excluded one class from the formal analysis. Most people approach sentence combining from a base in transformational-generative grammar (the source of its reinvention by English and Language Arts teachers in the 60s), but my theoretical base is more that of Vygotsky's (1978) zone of proximal development. I view sentence combining in ESL, especially with adult students, as a means of helping them transfer to L_2 competencies that they already control cognitively in the native language. One group was so definitely inferior to the other two that I felt that the material to be studied was truly beyond their zone of proximal development and that their inclusion in the control data would prejudice the results in favor of the experimental group. Since these are real classes with many variables beyond my control, my solution has been to prejudice everything possible in favor of the control group. I did, however, make all the collections from this third group, finding exactly what had been predicted.

The writing data consisted of a sentence combining exercise of paragraph length on a low-level scientific topic and a free writing sample on a daily conversation topic, assigned to be of paragraph length, with collections in the first and last weeks of the session. These were analyzed by the T-Unit indices, adding the error-free T-Unit measures as proposed by Larsen-Freeman and Strom (1977), except that spelling and punctuation errors were not counted.

During theseven weeks between the data collections, the groups studied the same grammar topics. The control group did exercises from the text and some compositions, also receiving instruction in outlining. Class discussion consisted of identifying and correcting errors. The experimental group began each grammar topic with similar types of exercises, first doing one or more orally before writing them. Next followed either a paragraph free writing assignment designed to elicit use of the structure under consideration or an open paragraph-or-longer sentence combining exercise calling for use of a variety of structures. The order of these varied. Then they were given an essay assignment, also designed to elicit the given structure. Occasionally they were allowed to select the topic for a composition. Selected paragraphs from the free writing assignments and the longer sentence combining exercises were corrected for error and displayed for class discussion as to how to make them "better." The students saw no errors except those they wrote themselves. All error correction was oral, frequently done by one student when another proposed a new

combination. The discussion focused on better ways of expressing the information contained in the paragraph. In most cases there was general agreement as to which of several options did the best job; occasionally the instructor would explain her decided preference for one solution.

These studies, then, are not comparing sentence combining with no sentence combining--sentence combining is traditional in our ESL program. Rather, this is a comparison of a-rhetorical, mechanical, directed, single-sentence-product sentence combining with the open, paragraph-level type described, in which there was a great deal of class discussion (two or three periods a week) of various ways to convey the same information.

T-Unit analysis of the pretest writing samples showed the two groups to be roughly equivalent. The control group scored non-significantly higher on most measures on both tasks. The only significant differences were in cl/T on the sentence combining task, in which the controls scored significantly higher because of greater use of reduction by the experimental group, and absolute length on the free writing task, in which the experimental students wrote significantly more words. On a free writing task, length is a factor that has to be interpreted; it is not necessarily good or bad <u>per se</u>.

I. MEAN SCORES ON PRETEST: SENTENCE COMBINING EXERCISE AND FREE WRITING (

	EXPERIMENTAL		CONTROL	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.
WORDS	104.08	15.83	115.57	23.0
	(131.0)	(55.94)	(76.18)	(23.37)
WD/T	12.08	2.97	13.06	3.67
	(11.78)	(2.72)	(14.78)	(6.07)
WD/CL	9.46	1.82	8.63	1.18
	(7.91)	(1.73)	(8.62)	(1.33)
CL/T	1.27	.17	1.52	.34
	(1.35)	(.27)	(1.78)	(.77)
%EFT	.50	.21	.64	.17
	(.47)	(.16)	(.39)	(.24)
%EFcl	.62	.21	.74	.13
	(.61)	(.12)	(.56)	(.21)
WD/EFT	9.9	5.72	11.32	3.78
	(10.21)	(3.1)	(10.56)	(4.9)

II. T-SCORES: EXPERIMENTAL COMPARED TO CONTROL

	SENTENCE COMBINING	FREE WRITING
WORDS	-1.5	+3.52*
WD/T	-1.29	-1.44
WD/CL	+1.3	-0.56
CL/T	-2.27*	-1.22
%EFT	-2.0	+0.75
%EFcl	-1.7	+1.0
WD/EFT	-1.4	+1.64

* Significant at .05

These charts show class pictures very like the instructors' impressions of the groups. Since one of the interests was to see if T-Unit analysis would be useful to us in placement/grading and another was to see if an exercise could be used instead of an essay, it was very encouraging to see that these objective, easily-made counts reflected our more subjective holistic impressions. Both groups were deemed "normal" by the instructors, both experienced in the level and in these particular groups. The control group was insignificantly better overall, but it contained a wider range of proficiency. This is normal, also.

Analysis of the posttest data shows a significant reversal of the relative position of the two groups, with the experimental group scoring higher on all error-free measures, significantly higher on two on each task, with a significant decrease in length on both tasks. The control group showed significant losses as a group, with only one student showing a significant error-free gain on the sentence combining task (wd/EFT) and three overall significant gains on error-free measures on the free writing task, all three having had at least two zeros in the three measures initially.

III. MEAN SCORES ON POSTTEST: SENTENCE COMBINING EXERCISE AND FREE WRITING()

	EXPERIMENTAL		CONTROL	
	X	S.D.	X	S.D.
WORDS	107.28 (96.14)	9.54 (32.83)	114.43 (106.79)	5.46 (33.68)
WD/T	12.7 (12.35)	2.2 (2.71)	12.76 (11.09)	3.52 (1.42)
WD/CL	7.87 (9.81)	1.14 (3.13)	6.94 (8.34)	.64 (1.11)
CL/T	1.67 (1.3)	.31 (.22)	1.82 (1.34)	.44 (.17)
%EFT	.58 (.90)	.25 (.10)	.42 (.52)	.24 (.20)
%EFcl	.74 (.92)	.17 (.08)	.59 (.58)	.19 (.19)
WD/EFT	12.61 (12.60)	4.11 (3.17)	7.86 (9.16)	2.19 (2.29)

IV. T-SCORES: EXPERIMENTAL COMPARED TO CONTROL

	SENTENCE COMBINING	FREE WRITING
WORDS	-2.37*	-0.80
WD/T	-0.05	-1.54
WD/CL	+0.25	+1.65
CL/T	-1.0	-0.57
%EFT	+1.68	+6.33*
%EFcl	+2.14*	+5.67
WD/EFT	+3.7*	+1.41

*Significant at .05

These data indicate that the type of sentence combining practiced by the experimental group resulted not only in increased syntactic maturity but also in decrease of error even though error was so rarely discussed in class. They show that the experimental group made positive changes whereas the control group actually sustained losses, particularly in the error-free measures, as they attempted to use the complex structures which they had studied but not mastered.

Error is disposed of in various ways (mostly ignored) in native speaker research. The error free counts are necessary in ESL. At this level students generally fall into two groups: those who write little strings of simple sentences strung together with "and", "but" and "so", and those who attempt to express themselves in English in the adult manner they would use in their own languages, in complex sentences replete with error. Increasing the repertoire of the first group is of little use if it merely places them in the second group. Both groups need instruction which will help them to express themselves correctly in an adult manner. This study indicates that open sentence combining practice with group discussion is an effective technique and that the "traditional" method is inadequate in an 8 week course. It is apparent from the figures that the control group was caught in a stage of getting worse in order to get better whereas the experimental group had begun to consolidate its gains. The 16-week projects should shed some light on the time factor.

These results are particularly revealing in view of the fact that the posttest exercise was the "Aluminum" passage, not particularly good for testing the overall proficiency of ESL students because of its heavy dependence on adjectival embedding, but excellent for this point in the project because a substantial portion of the material used in the control class dealt with the structures needed for this exercise. The experimental class spent only three days on formal study of adjective clauses (under the tutelage of an intern teacher) prior to the test. This passage should represent another bias in favor of the controls, both in content and in grammar.

Individual scores in the experimental group were then examined in exploration of Haswell's (1981) findings that above-mean experimental students wrote slightly shorter clauses than their control

counterparts and that most of the gain was found in the pretest below-mean students, a finding similar to that of Stedman (1971), whose black students showed the greater gains. The question here was, "Is there a type of student who benefits from this kind of sentence combining and another who does not?" This study provides no support for Haswell's conclusions; on the contrary, it indicates that the treatment might be considered a cure-all if subsequent research confirms the findings. Neither can it be concluded that length and proportion of EFTs provide an adequate, reliable index to quality in individual cases as proposed by Larsen-Freeman (1978). Three of the best writers in the experimental group began with problems of too many words. One wrote long Germanic sentences, far too packed with information for the American reader. One wrote "flowery" prose not suitable to the academic style being taught. One displayed such excessive adjectival embedding that I felt compelled to inquire if he had ever had a course in sentence combining or transformational grammar. For these three different reasons, all three students needed to decrease T-Unit length or change the style in some way that would make the length not an obvious problem. All three met the goal, all having decreases (two significant) in wd/T but increases by two in wd/EFT on the free writing task and significant increases on %EFT on both tasks.

These same students form part of the argument against Haswell's conclusion that sentence combining benefits the worst most (although carried to an extreme this would have to be true--if the student is an excellent writer, what technique would produce any observable benefits?). In this group, the above-mean experimental students wrote shorter T-Units but longer clauses than the control above-mean students, and the list of above-mean students is not identical either to the list of best writers or to the list of those who made the smallest positive changes. In the experimental group, it is clear that the students considered the balance of maximum information and maximum clarity, the instructor's principal contribution to the class discussions, and that they went beyond the principle of "longer is better" to use substantial reduction and a variety of sentence types. The good writers got better, not worse. The experimental sentence combining offered different things to each student, and it seems to have met the needs of all, including the student goal of passing the TOEFL, despite the lack of practice in error identification. In the control group, the best students generally showed the largest losses, and they were spread from bottom to top of the range of initial and final error-free measures. The small sample size precludes definitive statements in most areas, but it is clear that the use of two EFT measures from one writing sample would have resulted in improper placement for many individuals. Since the control curriculum was very like that of many native speaker sentence combining projects, Haswell's findings cannot be generalized.

The question of length is complex. In a sentence combining exercise, one can say that, in general, the shorter it is, the better because all students are dealing with the same bits of information. In any kind of free writing, however, the situation is less clear-cut. The final free-writing sample supported Larsen-Freeman's (1978) view that the more proficient writers adhered more closely to the assigned length. Because of their nature, these samples were not subjected to holistic analysis by outside raters; however, the investigator's judgments and objective scoring both place the better writers below the mean on absolute length,

with a significant improvement in both groups. The worse writers wrote far more than the paragraph assigned, with young students and Orientals writing the longest. In the 16-week projects, the free writing samples are assigned as "200-word essays", with an entire period allowed for composition and revision; these are being rated holistically by ESL members of the English Department. Only one rater has returned her scores, so the data have not been correlated. It is quite clear to the naked eye that her ratings will correlate highly with length, with content playing a much greater role than adherence to the assigned length or amount of error. It does not appear that there will be a strong correlation between the syntactic maturity scores and her ratings. (Incidentally, there are two native speakers in the current experimental group, both with language problems due to deafness. One was sent by his previous teacher, the rater mentioned above, who said that he had basic grammar problems. The second enrolled herself after hearing about it from him. She had the same teacher, but was considered a much better writer. My evaluation is that her language problems are far greater than his and that she is considered a better writer because she has more to say and develops her ideas better. I predict that she will show greater gains than he will because she is smarter. It is unfortunate that we do not have IQ data on our students.)

The age factor definitely should be investigated further. The evidence available so far supports the initial hypothesis that the younger students would benefit from a sentence combining course in Spanish first. In general they did the single-sentence-product exercises with skill and ease, but failed to see the need to combine in many obvious instances in the longer exercises and in free writing. The type of sentence and the type of paragraph wanted by the instructor and judged best by the older students may have been beyond their true zone of proximal development even though they made significant gains in both groups, the largest of all in %EFT.

The question of the influence of native language also remains to be investigated; it will take repeated studies to have sufficient evidence about specific languages to justify correlations. Larkin (1978) has proposed that Cantonese speakers have special difficulty with the adjective clause. That is contradicted by the small amount of evidence here. When the third project is completed, there will be enough subjects from several languages to permit looking at possible correlations.

The most dramatic finding of the 8-week study is that there were no article errors in the final free writing samples of the Oriental students in the experimental group; in fact, such errors essentially disappeared in about the sixth week. There was no discussion of articles in class except an occasional remark or question by a student in discussion of a different combination proposed for a passage. Preposition errors, also not discussed, became extremely rare. These two problems are notably resistant to frontal assault.

There was a pair of identical twins in this study, unfortunately both in the experimental group and neither willing to transfer to the other class. In the beginning, one was one letter grade better than the other even though both always said essentially the same thing in free and structured writing. It was possible to tell whose work was whose by looking at the syntax attempted. Both made significant changes,

but not in the same way or of the same magnitude, and at the end it was not possible to tell who had written which paper. Their improvement was primarily in decrease of error and absolute length. The one who began better had greater gains in syntactic maturity measures on the sentence combining task, but on the free writing task their changes are nearly identical or in favor of the initially worse one. One set of twins does not prove anything, but it leads me to believe that both progressed about as far as their very immature seventeen years permitted, with very large improvement in error-free measures as a result of their intelligence and liking for this method of instruction. Incidentally, as almost everyone who has reported on student attitude has found, ESL students also find this the most palatable way to learn grammar.

A major concern left in the area of writing analysis is what, if any, punctuation errors to count. In the data presented here, none were counted (making it impossible to compare directly with Larsen-Freeman's data). Given the lack of agreement among writers in the field, counting punctuation errors would introduce a rater variable that is now minimal. If punctuation had been counted in these studies, all error-free measures on the initial samples would have been at or near zero, with absurd change scores for the experimental students, who learned the punctuation as they mastered the complex structures. In the control group, the punctuation got fancier, but worse--there was much more of it, but it was highly incorrect. At this level, we do not expect them to punctuate correctly at entry. If I were to use punctuation also as a placement measure, I would be primarily concerned with three types of punctuation: putting something at the end of a T-Unit, putting a comma after a long initial adverbial or adjectival, and setting off "interrupters" (non-restrictive modifiers, conjunctive adverbs, etc.) with some form of punctuation. These seem to me to be the most correlated with syntactic understanding and with breathing; much other punctuation in English seems rather arbitrary and more variable from expert to expert.

Very little analysis has been done on the first 16-week project. We are awaiting the holistic ratings. What we have shows two very interesting things.

First, although their scores were comparable to those of the control group, the experimental class was judged by all who came in contact with them as being definitely not "normal." Everyone who taught them found it practically impossible to get them to talk about anything under any circumstances in the group. The class discussions did not go according to the plan--they consisted largely of question and answer, much prodding by the teacher. It was not surprising to find that the change scores after eight weeks were far from being as exciting as those in the previous project. The experimental class still made greater gains than the control class and received higher holistic ratings on the second composition than the controls did, but it appears that the discussion is a critical factor in producing the spectacular gains found in the first study. It <u>was</u> surprising that this Sphinx-like group also rated this activity very highly.

In the 16-week study, there was a third essay collection at the end. The results of this were quite a surprise. It appears that the topic was a very poor choice, far too vague and open, possibly causing the students to spend too much time deciding what to say in response to it. Both groups rated lower holistically on the third essay than on the second

one, with the control group falling even below its initial holistic scores. The T-Unit measures for both groups are very erratic, showing losses for both classes on all measures except wd/EFT in the control class. This does not match either instructor's picture of his class. In the current project, this topic will be assigned again, but a fourth will be added at about the same time, in the hope of proving that the fault lies with the topic.

There is little that can be said about the reading hypothesis until the computer results are available. The experimental class does show a significantly greater gain in reading scores than the controls, but that was an expected outcome based on several semesters of comparing the reading scores. The gain in the experimental class is insignificantly higher than previous reading classes taught by this instructor. The Propaganda Game, the principal teaching method in this class, is the more logical cause of the difference between the two classes. Future studies may help to isolate this factor as the present teacher of the writing control group intends to change to this type of sentence combining at the end of the current project.

Another problem with the reading hypothesis is the eternal problem of how to measure reading comprehension. Other sentence combining studies with a reading hypothesis have found positive correlation, negative correlation, and, most often, no correlation or insignificant gains. Investigator-made cloze tests have been most satisfactory in establishing positive correlation. The tests used in these studies are ESL reading materials (Smithies, 1972), which we had been using to test the effects of the Propaganda Game, again biasing the results in favor of the control class as much as possible because of the many variables in the reading classes and total program. It appears from inspection that the greatest gains in reading scores will correlate with those students' syntactic maturity gains, but the reverse will not be true because the student who scores high initially in reading cannot make a large gain on the posttest. It may, however, be useful to investigate an alternative hypothesis with these data. Hofmann (1979) has proposed that the influence is in the opposite direction, that the reading score is a predictor of the syntactic maturity score gain.

In June, 1981, I began a simple 8-week project, mainly to answer a few of these questions in ESL terms. Nine months later, the questions have multiplied like rabbits. I do not agree with John Mellon.

REFERENCES

Bateman, D. and F. Zisonis. 1966. The effect of a study of transformational grammar on the writing of ninth and tenth graders. Urbana, Ill.: National Council of Teachers of English.

Cooper, Thomas C. 1976. Measuring written syntactic patterns of second language learners of German. Journal of Educational Research 69:5, 176-183.

_____. 1978. Developing syntactic fluency of college foreign language students through sentence-combining practice. ERIC #166991.

_____. 1980. Sentence combining in second language instruction. ERIC #195167.

Daiker, Donald A., Andrew Kerek, and Max Morenberg. 1979. Sentence combining and the teaching of writing. Conway, Ark.: University of Central Arkansas.

Frank, Marcella. 1972. Modern English, part II: sentences and complex structures. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Haswell, Richard H. 1981. Within-group distribution of syntactic gain through practice in sentence-combining. Research in the Teaching of English 15:1, 87-96.

Hofmann, Richard J. 1979. The relationship of reading comprehension to syntactic maturity and writing effectiveness. In Daiker, Kerek, and Morenberg, 109-115.

Hunt, Kellogg. 1965. Grammatical structures written at three grade levels. Champaign, Ill.: National Council of Teachers of English.

Larkin, Greg and Ron Shook. 1978. Interlanguage, the monitor, and sentence combining. ERIC #169779.

Larsen-Freeman, Diane and Virginia Strom. 1977. The construction of a second language acquisition index of development. Language Learning 27:1, 123-134.

Larsen-Freeman, Diane. 1978. An ESL index of development. TESOL Quarterly 12:4, 439-448.

Mellon, John. 1979. Issues in the theory and practice of sentence combining: a twenty-year perspective. In Daiker, Kerek, and Morenberg, 1-38.

Smithies, Michael. 1972. Advanced English Comprehension texts for science students. London: Collier-Macmillan Publishers.

Vygotsky, Lev S. 1978. Mind in society. Cambridge, Mass.: Harvard University Press.

Zamel, Vivian. 1980. Re-evaluating sentence-combining practice. TESOL Quarterly 14:1, 81-90.